AD_____

Award Number: DAMD17-01-1-0559

TITLE: Genome Wide Expression Analysis of Breast Cancer in African American

PRINCIPAL INVESTIGATOR: Lisa B. Baumbach, Ph.D.

CONTRACTING ORGANIZATION: University of Miami
Miami, Florida 33101

REPORT DATE: October 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2004	Annual (1 Sep 2001 - 1 Sep 2003)

4. TITLE AND SUBTITLE
Genome Wide Expression Analysis of Breast Cancer in African American

5. FUNDING NUMBERS
DAMD17-01-1-0559

6. AUTHOR(S)
Lisa B. Baumbach, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Miami
Miami, Florida 33101

E-Mail: l.baumbach@miami.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)
No abstract provided.

14. SUBJECT TERMS
No subject terms provided.

15. NUMBER OF PAGES
9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

February 10, 2005

Shelly Marken, Grants Manager
Congressionally Directed Medical Research Programs

Judy Pawlus, Technical Editor, USAMRMC

RE: Award No. DAMD17-071-0559 PROGRESS REPORT
"Genome Wide Expression Analysis of Breast Cancer in African Americans"

Dear Ms. Marken and Pawlus,

Please let this letter serve as an interim progress report for the above referenced grant. I would like to start with the most recent events, and work my way in reverse chronological order.

Brief Overview of Project

This project has been focused on the preliminary investigation of whole genome changes associated with breast cancer in AA women This project was funded as a Concept Award from the Department of Defense in 2000. The original aims were to analyze total RNA from tissue that would be surgically removed from breast cancer tissue and non-cancerous breast tissue of three AA women with a positive family history of Breast Cancer, and three women without a family history. RNA extraction, experimental design, hybridization experiments, and data analyses are completed in conjunction with the DNA microarray facility at UM Medical School. Unfortunately, it took a very long time to achieve Human Subjects approval from both the Department of Defense and the Human Subjects committee at the medical school, (as documented in previous letters) as well as to get agreement on the patient recruitment criteria and the tissue procurement protocol by the consulting surgeons and pathologists. Finally, in June 2004, we enrolled our first subject into the study, and despite setbacks from numerous hurricanes in Summer, 2004, we have four independent cases enrolled (two with strong family histories of breast cancer), cancerous and self-matched normal tissue obtained, and RNA samples ready to be hybridized to Whole Genome DNA microarrays in January, 2005.

Recent Updates

In the last two weeks, we have received the good news that our microarray experiments have worked! We now have a very large data set to analyze, which will take several months to complete the statistical analyses. We do know that we have a list of genes that are over- or under- expressed two or more times, and we are in the process of organizing them into biological pathways. All of these experiments were completed in house- at the University of Miami Microarray facility in collaboration with my laboratory. They are the very first human microarray experiments to be done at our institution!

These experiments were performed using the *Agilent Whole Human Genome* arrays. This single 44K formatted microarray represents a compiled view of the human genome as it is understood today. The sequence information used to design this product was derived from a broad survey of well-known sources such as *RefSeq, Goldenpath, Ensembl, Unigene* and others. The resulting view of the human genome covers 41K unique genes and transcripts that have been verified and optimized by alignment to the human genome assembly and by *Agilent's Empirical Validation process* (http://www.chem.agilent.com/). Total RNA was extracted from excised tumor and non-cancerous (self-matched) breast tissue from the patients listed below. Enough RNA was purified that we were able to use these samples directly in hybridization experiments.

African-American samples for Microarray Analysis 2004

Sample ID	Current Age	Date Obtained	Tumor Tissue	Normal Tissue	Family History	Summary	Status
BM0401	45 yrs	6/16/04	+	+	Maternal Aunt BC	Partial mastectomy/ invasive ductal carcinoma	RNA extracted/ Hybridization successful
BM0402	50 yrs	7/6/04	+	+	Sister with BC	Right breast infiltrating adenocarcinoma	RNA extracted/ Hybridization successful
BM0403	56 yrs	10/7/04	+	+	Possible Sporadic/ confirming Family History	Left breast lumpectomy/ axillary node dissection	RNA extracted/ Hybridization successful
BM0404	55 yrs	11/24/04	+	+	Mother & Grandmother with BC	Left breast mastectomy	RNA extracted/ Hybridization successful

Tumor RNA was labeled with Cy3 fluorescent dye; normal RNA was labeled with Cy5 fluorescent dye. The two RNA samples (per patient) are equally mixed, and hybridized to the *Agilent Whole Human Genome* arrays. In order to ensure internal quality control, the samples have a dye swap also done (i.e., Tumor RNA – Cy5 and Normal RNA- Cy3), and the two data sets compared for quality and internal consistency.

The first pass data analysis is carried out with the *GenePix* software and provides image processing capabilities to quantitate the Cy3 and Cy5 fluorescence signals from each array. Briefly, a grid is superimposed on the array area according to the printing pattern of the array. This grid is aligned with the underlying image and comparing the fluorescence signal (at each wavelength) in the grid to the background then identifies features. Various statistics are computed for each wavelength and the data are exported from *GenePix* in a 'Results' file (*http://www.spotfire.com*.).

This file computes more than 40 characteristics of each feature on the array including # pixels in the feature, # pixels in the surrounding background, size of the feature (can be fixed or fitted during analysis), etc. This is the data set that we currently have.

The results then need to be analyzed using the following programs: *Spotfire Pro/Array Explorer (Spotfire Inc.,* Cambridge, MA) and *Javelin* (Axon Instruments, Foster City, CA). *Spotfire* is used as the initial data visualization tool, and as a clustering tool to identify groups of genes that show similar expression profiles. Hierarchical cluster analysis is done according to Eisen and colleagues (*http:www.microarrays.org/ software.html*). This approach allows relatedness of genes to be tested. In this manner, whether gene expression is increased or decreased, results will be analyzed by a "filtered" approach. The data set is then moved to *Javelin*, a newly developed analysis suite with an SQL-compliant database. *Javelin* provides the network-based database in which all expression data is stored.. One of the key strengths of this software is the range of data normalization tools provided. These tools enable us to quickly analyze array data in a variety of ways to determine the best normalization approach for a given experiment. This is particularly important for experiments that analyze data sets from many different RNA isolations, as described in this application. Expression data from micoarray assays will become part of our institution's expression database. Currently, this database is for use by institutional investigators and their collaborators only, but should be moved shortly into the public domain (accessible through our institutional Web Site).

Data Analysis and Interpretation of Results: Results of all experiments will need to be analyzed and compared at several levels. Results will need to be compared both within groups and across groups (familial and sporadic) for similarities and differences. It is anticipated that (1) some common differences in gene expression will be detected in the cancerous tissue analyzed, and (2) certain gene expression differences will exist. We will also compare results to *BRCA1/BRCA2* mutation positive and negative subjects, as well as to already published literature (Hedenfalk et.al. 2001; Segal et.al. 2004), and very recent reports we will discover through *PubMed* searches.

Highlights of Past Efforts:

June, 2004-present- Collection of patient samples for studies; RNA isolation; refinement of study design for microarray experiments; hybridizations completed for four patients.

November, 2003-June, 2004- Design tissue collection strategies; refine patient enrollement criteria, meet numerous times with breast surgeons and pathologists to develop actual implemented tissue collection protocol; work with UM microarray facility to refine study design, and to decide on actually DNA microarrays to use.

August, 2003-November, 2003- Extensive review of the literature concerning published breast cancer microarray studies (see attached table to this report), review and analysis of these studies. Came to conclusion that (1) no microarray studies had been published for African-Americans; (2) no common microarray platform existed between all of these

studies, therefore, we could use the best available platform at our institution; (3) the Agilent 40K Human Genome microarrays would soon be available for use, and we should wait to use these, as they were state-of-the art.

June, 2003-July, 2003- Tried to work out between UM and FAU (in Boca Raton) and way to use Affymetrix arrays for the microarray studies- plan never worked, and we had to return to the drawing board regarding source of microarray to use for experiments. Decided to conduct extensive literature review to assess current state of the field.

May, 2003- Final approval of Human Subjects approval granted from UM. Letter sent to DOD requesting one year no-cost extension to project. Significant difficulties in obtaining final IRB approval (from late 2002 until May, 2003).

September, 2002- Letter sent to DOD requesting one year no-cost extension to project, due to initial problems with experimental performance was that we had proposed to use new DNA microarray technology to analyze genome wide mRNA expression in breast cancer tissue from African American women. In early fall of 2001, when the initial award period began, our medical school was negotiating with *Incyte* corporation to develop and purchase "start-up" packages for investigators conducting projects using microarray technology, which would provide all the necessary protocols, reagents, equipment, and software packages for data analysis, so my laboratory had agreed to this, and was basically ready to begin the proposed experiments. However, in late October, *Incyte* executives made a totally unexpected *carte blanche* decision to discontinue all services relayed to microarray production and development, and thus, this decision obliterated our start-up package agreement. This left us with no other alternatives at that time. In September, 2002, we were also in the midst of the on-going negotiation and review between the Department of Defense (DOD) , our institution, and the P.I. regarding the Human Subject's protocol and consent documents

Literature Cited

Hedenfalk I, Duggan D, Chen Y *et al.* Gene-expression profiles in hereditary breast cancer. *New Engl J Medicine* 2001, 344:539-548.

Segal E, Friedman N, Koller D *et al.* A module map showing conditional activity of expression modules in cancer. *Nature Genet.* 2004, 36:1090-1098.

I hope this letter suffices as an interim progress report. Please let me know if you need additional material.

Sincerely,

Lisa L. Baumbach, Ph.D., P.I.,
Associate Research Professor
Department of Pediatrics

Summary of Breast Cancer Microarray Literature Review

References	Sample Type	Type of Microarray	Results
1.Berns, 2001	Frozen, 6 with BRCA1 (7?), 12 sporadic matched for 9 criteria plus one cell line. F samples age 30-60, sporadic 20-50, mostly idc tumors, ? no mention tumor grade. Dutch ethnic	Clontech Atlas, Cancer cDNA filters with 588 genes. Hierarchical clustering (Eisen Stanford). Log transformed TreeView 340/588 in final analysis. Spearman Rank correlation, sample size too small for stats	Two major groups: A: Tumors node negative, fewer TP53 gene alterations increased genes cytokeratin 14 BAX, CyclinD1 EGFR B: 5/6 BRCA1 tumors (6^{th} was 2^{nd} after treatment)
2.Hedenfalk,I NEJM,2001,344	Flash frozen 7 with BRCA1, 8 with BRCA2, 7 sporadic. BRCA1 grade 3. Patients from Sweden. BRCA test information ref 18	CDNA array, sequences from Res Genetics (also tissue array, Swedish pop. Series). Class prediction method, gene list generated, modified 1. F/T test 2. weighed gene analysis 3. Info Score, "mutal info scoring", following with agglomerative hierarchical clustering	They found 3 different gene expression profile groups: BRCA1, BRCA2 and sporadic. 176 genes were differentially expressed in BRCA1 and BRCA2 tumors
3.Sorlie, T. PNAS,2001	78 breast ca. tumors from 77 different individuals,4 normal breast tissues as reference. ER, PR & TP53 checked, for clinical data see table 1, in www.pnas.org	8,102 genes per chips (Stanford Univ. protocol), analyzed by hierarchical clustering and by SAM algorithm. Searched for genes that correlated with survival.	They could be able to identified 5 distinct groups with different survival outcomes.
4.Welsh,JB. PNAS,2001,98	27 papillary adenocarcinomas of the ovary and 3 normal samples of ovarian tissue used as reference. For clinical data see table in www.pnas.org	Oligonuceotide arrays complementary to more than 6000 human genes. They compared the expression of normal and malignant tissues (protocol as ref.9).Image analysis by genechip 3.1 Affimetrix. Expression variation was processed by Cluster and Treeview (ref 12).	They selected 1243 genes with strongly varied expression (SD>250). By hierarchical clustering they found different expression patterns for normal and malignant tissues and the last one could be subdivided into mayor groups with histological and clinical correlations (www.gnf.org./cancer/ovary)
5.Van't Veer LJ. Nature,2002,415	98 br. canc.:34 with metastases & 44 cancer free within 5 years, 18 with BRCA1 and 2 with BRCA2 mut. All "sporadic" (no BRCA mut) patients were lymph node negative and <55years.	2 cRNA hybrid1zations per tumor with 25000 genes arrays synthesized by inkjet technology Expression was taken as the intensity ratio between target and a reference pool (pooled equal amounts of cRNA from each of the sporadic carcinoma(ref 14). 5000 different expressed genes were selected (2fold difference) by unsupervised hierarchical clustering algorithm.	After correlation coefficient between disease outcome and gene expression was applied they found 231 genes significantly associated with disease outcome; these groups of genes were rank ordered on the basis of the correlation coefficient magnitude, then a prognosis classifier was optimized by sequentially adding subset of 5 genes from top, at the end, the optimal number was 70 genes. The classifier predicted correctly in 83% of the patients. The poor prognostic signature consist of genes from cell cycle, invasion, metastases and angiogenesis.

6. Jazaeri,AA. JNCI,2002,94	61 patients with ovarian cancer (18 linked to BRCA1, 16 to BRCA2 and 27 sporadic), all were askenazi jews. Reference for cDNA Microarray was pooled RNA from virus 40 immortalized cell lines	cDNA microaaray schips with 7651 total features manufactured at NCI Microarr. facility. Data was normalized and filtered, modified F test applied (P,0.0001), also multidimensional scaling and hierarchical clustering applied Results checked by semi quantitative rt-pcr in 6 genes from 5 tumor samples	They selected 110 genes with strong differences in expression and with them they could classified the tumors in a BRCA1 pattern, a BRCA2 pattern and inside the sporadic tumors in a BRCA1like and a BRCA2like patterns.
7. Jiang, y. et al 2002, Oncogene 21:2270-2282	3 1° breast tumors for subtracted cDNA library. For arrays: 22 1° breast tumors, 3 tumors meta. to lymph nodes, 2 normal breast samples, 30 samples variety of normal essential tissues	cDNA array 2000 randomly selected cDNA clones generated by subtraction experiment. Arrays made with Incyte chemistry. 62 of these with $\geq$ threefold expression were used in further analysis.	62 clones with 3x expression difference were analyzed resulting in four genes which appear to be breast tumor specific: B709P, B736P, hair specific type ii keratin, a GABA subunit.
8. Sotiriou CH, PNAS,2003,100	99 tumor samples (invasive ductal carcinomas): 46 node neg. and 53 node positive, all patients underwent radio, chemo and hormonal therapy. Total RNA from Univ. Human Refer. (Sratagene) was amplified and used as reference.	CDNA chips with 7650 total features from NCI (protocol at http://nciarray.nci. nih.gov/reference/ Index.shtml It was compared espression profiles between specimens classified according to standard prognostic classific. (EX. Tumor grade 1,2,3 node-and+, etc). Cluster analysis was applied over 706 genes selected for expression variation	Table1 shows the number of genes significantly expressed according to clinical-pathological parameters like ER neg. vs. and ER positive, tumor grade, etc. They concluded that ER status was the best correlated with gene expression.
9. Hedenfalk,I. PNAS, 2003,100	Familial cases with no BRCA1/2 mutations: BRCAx patients (8 families, 2-3 tumor per family from different individuals) In summary 16 tumors from 8 BRCAx families	6500 cDNA clones representing 4700 unique known genes and 1700 ESTs. Breast cancer cell lines (ATCC) was used as reference. Hierarchical clustering and multidimensional scaling analysis was applied	They identified 2 different groups of tumors based on 60 genes, many of them were ribosomal genes. The CYPIAI gene was over expressed.
10. Zhu, G. Oncogene, 2003,22	10 breast cancer tumors: 5 infiltrating adenocarcinomas and 5 Ductal Carcinoma In Situ (DCIS). In these tumors rim cells were compared with central cells (both taken by microlaser dissection). To be compared, all samples from the different tumors were pooled together in two aliquots: "pooled peripheral" and "pooled central"	Human Cancer 1.2 Arrays (Clontech, Palo Alto, Ca) nylon arrays, 1176 known cancer associated genes (cell adhesion, cell cycle regulation, tumor suppressor genes, including negative control and housekeeping genes) were used, see http://www.clontech .com (Atlas/genelist) Each microarray hybridization used a pool of 5 samples. 2 genes were selected for validation with rt-pcr	From the 1176 genes analyzed 9 genes were up regulated and 3 genes down regulated in the periphery of DCIS relative to the central area. In the frankly invasive tumors 5 were up regulated and 4 down regulated. Thus, changes in gene expression associated with variation in micro anatomical location of neoplastic cells can be detected within even small developing tumor masses

11. Dressman, MA. Cancer Research,2003,63	34 breast tumor from Sweden patients: 14 ER neg., 19 ER positive, 1 ER ? see table 1 for clin-pathol classification	Affimetrix protocol. HuGene Fl (6800)?? They did not specified the type of arrays	ERBB2 was frequently over expressed predominantly in ER neg. tumors, 3 more genes of the same chromosomal region were also over expressed, what they also corroborated using gene copy studies
12. Huang, E. The Lancet, 2003,361	89 tumor samples see table 1 for clinical information, patients are from Taipei, China	Affimetrix chips U95AV2 with 12000 genes and ESTs, using affimetrix protocol and used 3 new statistical models (correlation-based clustering)	They found aggregate patterns of gene expressions associated with lymph node status and recurrence